(1) Approximately what percent of countries have GDP per capita between $20,000 and $30,000? **(D)**

(2) How would you describe the distribution of GDP per capita across countries? **(A)**

(3) Which of these could be the median of these data? **(A)**

(4) Data on the total compensation for a sample of CEOs is highly skewed. The variable recording compensation is named X and it is measured in dollars. Consider creating a new variable Y, which transforms X. Which of the following would result in a more symmetric histogram? **(C)**

(5) If you transform a variable using a linear transformation that adds a constant to all observations then which of the following would NOT change? **(B)**

(6) Which box plot could summarize these 2012 employment rate data? **(E)**

(7) In Galton's original research, which has observations for over 1,000 father-son pairs, the heights of sons are regressed on the heights of the fathers. Why would the height of mothers be an unobserved variable (i.e. a lurking or confounding variable)? **(B)**

(8) What kind of data are these? **(A)**

(9) If all of the observations except for the first one (Albania) were dropped, what kind of data would the remaining data be? **(B)**

(10) If all of the variables except for "inf11" were dropped (not counting the country names), what kind of data would the remaining data be? **(C)**

(11) Using these data, why would a cross-tabulation be a poor way to assess how inflation in 2000 is related with inflation in 2010? It would be a poor choice because a cross-tabulation ___. **(A)**

(12) For which year is the interquartile range of inflation the largest? **(B)**

(13) For which year is the range of inflation the largest? **(A)**

(14) In 2005 roughly what fraction of countries in these data had inflation below 5%? **(C)**

(15) This STATA summary of inflation corresponds to which year? **(A)**

(16) For two interval variables the coefficient of correlation is 0.12 and the s.d. of the residuals is 1.4. Which is NOT a reason to check a scatter diagram before calculating these statistics? **(C)**

(17) These data are ___. **(B)**

(18) Hall's regression analysis assumes a linear relationship between growth and ___. **(B)**

(19) If Gibrat's Law were true we would expect ___ between firm size and growth. **(B)**

(20) Hall finds heteroscedasticity. Which is a good explanation for why there is heteroscedasticity? **(B)**
(21) About what percent of the variation across the OECD countries in the percent of women employed in 2012 is explained by variation in the percent of women employed in 2006? **(C)**

(22) How do you interpret the slope of the OLS line? **(C)**

(23) If instead of percentage points the employment rates of females in both 2006 and 2012 were recorded as the fraction employed (e.g. 0.61 instead of 61%) what would the covariance be? **(A)**

(24) If you randomly select an undergraduate student from U of T what is the probability that s/he is an international student? **(B)**

(25) If you randomly select a student from the Scarborough campus what is the probability that s/he is an international student? **(D)**

(26) Consider three independent surveys of undergraduate students where a random sample of 1,000 students is selected for each one. What is the probability that the same student (Albert Chung) is selected for two out of the three surveys? **(D)**

(27) If domestic undergraduate students pay $8,000 of tuition and international undergraduate students pay $27,000 in tuition then how much do undergraduate students pay on average? **(E)**

(28) A survey of 1,000 undergraduates asks how much tuition the student pays. It yields an average that is less than the answer to the previous question. Which of these is a good explanation for why non-response bias may be the cause of part of the discrepancy? **(D)**

(29) This regression line is not an estimate of the supply curve. Why not? **(B)**

(30) When trying to estimate the supply curve using data on quantity produced and price, what would be a good example of an unobserved (i.e. a lurking or confounding) variable? **(E)**

(31) What kind of data does this research use? **(B)**

(32) The finding about the probability of dying is a comparison of ___ probabilities. **(C)**

(33) What percent of females are aged 65 to 69? **(C)**

(34) Focusing on people aged 30 to 34, what can we say about the proportion of males versus females? **(C)**

(35) What is the 95th percentile of age? **(A)**

(36) When is the slope of the regression line the same as the coefficient of correlation? **(C)**

(37) What percent of customers who visit the site make a purchase? **(D)**

(38) If you know that a visitor did NOT make a purchase, what is the chance that s/he had been offered nothing? **(E)**

(39) Consider a repeat of this experiment that finds that 10 percent of visitors make a purchase. If purchasing behavior were completely independent of any offers of discounts and free shipping then what would be the probability that a visitor offered nothing would make a purchase? **(B)**